A NOTE ON GENERAL AND SPECIFIC COMBINING ABILITY

BU-78-M W. T. Federer and C. R. Henderson July, 1956

I. Introduction

The expectation of the mean square for specific combining ability has appeared in several forms in the literature [1-7]. The question concerning the "correctness" of the various forms has arisen from time to time. All forms are correct, but different assumptions and definitions are involved in obtaining the various expectations. The purpose of this paper is

(i) to set forth the specific and general combining ability problem in detail,

(ii) to set forth the assumptions and definitions required to obtain the various expectations,

(iii) to discuss the validity of the assumptions and definitions, and

(iv) to redefine specific combining ability.

II. The Experimental Situation

The linear model usually assumed for the possible crosses among p lines observed in a completely randomized design is

$$Y_{ijh} = \mu + \alpha_i + \alpha_j + \delta_{ij} + \epsilon_{ijh} , \tag{II-1}$$

where Y_{ijh} represents an observation on the hth offspring (or group of off-springs) from the cross of line i with line j, μ = an effect common to all observations, α_i = the effect of the ith line, α_j = the effect of the jth line, δ_{ij} = an effect common to ith line crossed with jth line = nicking effect = effect due to dominance and epistasis, ϵ_{ijh} = random component associated with measurement of the effects. It is assumed that there are n offspring from each cross, that there are p(p-1)/2 crosses (i.e., there are no reciprocals), and that the ϵ_{ijh} are random independent variates with mean zero and common variance σ_{ϵ}^2.

The addition of other additive effects for stratification, such as blocks, does not affect the expectation of the mean squares for general and for specific combining ability. Hence, no effects for stratification are included in the model.

The breakdown of the degrees of freedom in the analysis of variance for the experiment described above is:

Source of variation	Degrees of freedom	Sum of squares	Mean square
Among lines	$p-1$	G	G'
Among crosses within lines	$\frac{p(p-3)}{2}$	S	S'
Among yields of the same cross	$\frac{p(p-1)(n-1)}{2}$	E	E'
Total	$\frac{np(p-1)}{2} - 1$	T	-

The normal equations for the completely randomized design with p(p-1)/2 crosses, each with n offspring (or with n groups (plots) of offspring) per cross, are obtained from the following table of totals, each from n observations:

line i

line j	1	2	3	4	5	... p
1	-	$Y_{12\cdot}$	$Y_{13\cdot}$	$Y_{14\cdot}$	$Y_{15\cdot}$	... $Y_{1p\cdot}$
2	-	-	$Y_{23\cdot}$	$Y_{24\cdot}$	$Y_{25\cdot}$	... $Y_{2p\cdot}$
3	-	-	-	$Y_{34\cdot}$	$Y_{35\cdot}$	... $Y_{3p\cdot}$
4	-	-	-	-	$Y_{45\cdot}$	... $Y_{4p\cdot}$
5	-	-	-	-	-	$Y_{5p\cdot}$
⋮						
p	-	-	-	-	-	-

The residual sum of squares is:

$$R = \sum_{i<j=2}^{p} \sum_{h=1}^{n} (Y_{ijh} - \mu - \alpha_i - \alpha_j - \delta_{ij})^2 ; \quad \text{(II-2)}$$

the partial derivatives with respect to μ, α_i, α_j, and δ_{ij} are:

$$\frac{dR}{d\mu} = -2 \sum_{i<j} \sum_{h} (Y_{ijh} - \mu - \alpha_i - \alpha_j - \delta_{ij}) ; \quad \text{(II-3)}$$

$$\frac{dR}{d\alpha_i} = -2 \sum_{j\neq i} \sum_{h} (Y_{ijh} - \mu - \alpha_i - \alpha_j - \delta_{ij}) ; \quad \text{(II-4)}$$

$$\frac{dR}{d\alpha_j} = -2 \sum_{i\neq j} \sum_{h} (Y_{ijh} - \mu - \alpha_i - \alpha_j - \delta_{ij}) ; \quad \text{(II-5)}$$

$$\frac{dR}{d\delta_{ij}} = -2 \sum_{h} (Y_{ijh} - \mu - \alpha_i - \alpha_j - \delta_{ij}) . \quad \text{(II-6)}$$

The various normal equations are:

<u>For the mean</u> $\hat{\mu}$

$$\frac{np(p-1)}{2} \hat{\mu} + n(p-1) \sum_{i=1}^{p} \hat{\alpha}_i + n \sum_{i<j} \hat{\delta}_{ij} = \sum_{i<j} \sum_{h} Y_{ijh} = Y_{\cdots} . \quad \text{(II-7)}$$

For the $\hat{\alpha}_i$

$$n(p-1)(\hat{\mu}+\hat{\alpha}_i) + n \sum_{j \neq i} (\hat{\alpha}_j + \hat{\delta}_{ij}) = \sum_{j \neq i} \sum_h Y_{ijh} = Y_{i..} \quad . \tag{II-8}$$

For the $\hat{\alpha}_j$

$$n(p-1)(\hat{\mu}+\hat{\alpha}_j) + n \sum_{i \neq j} (\hat{\alpha}_i + \hat{\delta}_{ij}) = \sum_{i \neq j} \sum_h Y_{ijh} = Y_{.j.} \quad . \tag{II-9}$$

For the $\hat{\delta}_{ij}$

$$n(\hat{\mu} + \hat{\alpha}_i + \hat{\alpha}_j + \hat{\delta}_{ij}) = Y_{ij.} \quad . \tag{II-10}$$

Using the following relations:

$$\sum_{i=1}^{p} \hat{\alpha}_i = \sum_{j=1}^{p} \hat{\alpha}_j = 0 \quad \text{and} \tag{II-11}$$

$$\sum_{i \neq j} \hat{\delta}_{ij} = \sum_{j \neq i} \hat{\delta}_{ij} = 0 \quad , \tag{II-12}$$

the following solutions for the estimates $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\alpha}_j$, and $\hat{\delta}_{ij}$ are obtained:

$$\hat{\mu} = 2\, Y_{...}/np(p-1) = \frac{1}{p} \Sigma \bar{y}_{i..} = \bar{y} \quad ; \tag{II-13}$$

$$\hat{\alpha}_i = \frac{1}{n(p-2)} \left\{ Y_{i..} - n(p-1)\hat{\mu} \right\} ; \tag{II-14}$$

$$\hat{\alpha}_j = \frac{1}{n(p-2)} \left\{ Y_{.j.} - n(p-1)\hat{\mu} \right\} ; \tag{II-15}$$

$$\hat{\delta}_{ij} = \frac{1}{n(p-2)} \left\{ (p-2)Y_{ij.} - Y_{i..} - Y_{.j.} + \frac{2Y_{...}}{p-1} \right\} . \tag{II-16}$$

The various sums of squares are:

<u>Total with $np(p-1)/2 - 1$ d.f.</u>

$$T = \sum_{i<j=2}^{p} \sum_{h=1}^{n} Y_{ijh}^2 - 2Y_{...}^2/np(p-1) \quad . \tag{II-17}$$

Among lines with p-1 d.f.

$$G = \sum_{i=1}^{p} \frac{Y_{i..}^{2}}{n(p-2)} - \frac{4\,Y_{...}^{2}}{np(p-2)}$$

$$= \frac{4}{np^{2}(p-2)} \ \Sigma \ \left(\frac{p}{2} Y_{i..} - Y_{...}\right)^{2} \quad . \qquad \text{(II-18)}$$

Among crosses within lines with p(p-3)/2 d.f.

$$S = \sum_{i<j} \frac{Y_{ij.}^{2}}{n} - \frac{\Sigma\, Y_{i..}^{2}}{n(p-2)} + \frac{2Y_{...}^{2}}{n(p-1)(p-2)}$$

$$= \sum_{i<j} Y_{ij.} \left\{ (p-2)\, Y_{ij.} - Y_{i..} - Y_{.j.} + \frac{2Y_{...}}{p-1} \right\} \Big/ n(p-2)$$

$$= \sum_{i}\sum_{\substack{j \\ i \neq j}} \left[(p-2)Y_{ij.} - Y_{i..} - Y_{.j.} + \frac{2Y_{...}}{p-1} \right] \Big/ n(p-2)^{2} \quad . \qquad \text{(II-19)}$$

Among yields within crosses with p(p-1)(n-1)/2 d.f.

$$E = \sum_{i<j} \left\{ \sum_{h} Y_{ijh}^{2} - Y_{ij.}^{2}/n \right\} \quad . \qquad \text{(II-20)}$$

III. Expectation of Mean Squares

III - 1. Random effects case

If it is assumed that the p lines represent a random sample of lines from a large population, then the expectation of the mean squares G' and S' are, respectively:

$$E[G'] = \frac{1}{p-1} E\left[\frac{\sum_{i=1}^{p} Y_{i..}^2}{n(p-2)} - \frac{4Y_{...}^2}{np(p-2)}\right]$$

$$= \frac{1}{np(p-1)(p-2)} \Big\{ p \sum_i E\,[n(p-1)(\mu+\alpha_i) + n \sum_{j\neq i} (\alpha_j+\delta_{ij}) + \sum_{j\neq i} \sum_h \epsilon_{ijh}]^2$$

$$- 4E\,[\frac{np(p-1)}{2}\ \mu + n(p-1) \sum_{i=1}^{p} \alpha_i + n \sum_{i<j} \delta_{ij}$$

$$+ \sum_{i<j} \sum_h \epsilon_{ijh}]^2 \Big\} = \frac{1}{np(p-1)(p-2)} \Big\{ p^2 [n^2(p-1)^2 (\mu^2+\sigma_\alpha^2) +$$

$$n^2(p-1)(\sigma_\alpha^2+\sigma_\delta^2) + n(p-1)\sigma_\epsilon^2] - [n^2p^2(p-1)^2\hat{\mu} + 4n^2p(p-1)^2\sigma_\alpha^2$$

$$+ 2n^2p(p-1)\sigma_\alpha^2 + 2np(p-1)\sigma_\epsilon^2] \Big\}$$

$$= \sigma_\epsilon^2 + n\sigma_\delta^2 + n(p-2)\sigma_\alpha^2 \ . \qquad \text{(III-1)}$$

$$E[S'] = \frac{2}{p(p-3)} E \left\{ \sum_{i<j} \frac{Y_{ij.}^2}{n} - \frac{\sum_{i=1}^{p} Y_{i..}^2}{n(p-2)} + \frac{2Y_{...}^2}{n(p-1)(p-2)} \right\}$$

$$= \frac{2}{np(p-2)(p-3)} \Big\{ (p-2 \sum_{i<j} E\,[n(\mu+\alpha_i+\alpha_j+\delta_{ij}) + \sum_h \epsilon_{ijh}]^2$$

$$- \sum_i [n(p-1)(\mu+\alpha_i) + n \sum_{j\neq i} (\alpha_j+\delta_{ij}) + \sum_{j\neq i} \sum_h \epsilon_{ijh}]^2$$

$$+ \frac{2}{p-1} E\,[\frac{np(p-1)}{2}\ \mu + n(p-1) \Sigma\alpha_i + n\sum_{i<j} \delta_{ij} + \sum_{i<j} \sum_h \epsilon_{ijh}]^2$$

$$= \frac{2}{np(p-2)(p-3)} \Big\{ (p-2)\frac{p(p-1)}{2}\ [n^2\mu^2 + 2n^2\sigma_\alpha^2 + n^2\sigma_\delta^2 + n\sigma_\epsilon^2]$$

$$- p\ [n^2(p-1)^2\mu^2 + [n^2(p-1)^2 + n^2(p-1)\ ]\ \sigma_\alpha^2 + n^2(p-1)\sigma_\delta^2 + n(p-1)\sigma_\epsilon^2]$$

$$+ \frac{2}{(p-1)} [\frac{n^2p^2(p-1)^2}{4} \mu^2 + n^2(p-1)^2p\sigma_\alpha^2 + n^2\frac{p(p-1)}{2}\sigma_\delta^2 + n\frac{p(p-1)}{2}\sigma_\epsilon^2]$$

$$= \sigma_\epsilon^2 + n\sigma_\delta^2 \ . \qquad \text{(III-2)}$$

In the above expectations it is assumed that the effects α_i, δ_{ij}, and ϵ_{ijh}, are independent random effects with mean zero, and that

$$E[\alpha_i^2] = \sigma_\alpha^2 \quad ; \tag{III-3}$$

$$E[\delta_{ij}^2] = \sigma_\delta^2 \quad ; \tag{III-4}$$

$$E[\epsilon_{ijh}^2] = \sigma_\epsilon \quad ; \tag{III-5}$$

$$E[\mu^2] = \mu^2 \quad ; \tag{III-6}$$

$$E[\epsilon_{ijh}\,\epsilon_{efg}] = 0 \quad \text{except where } ijh = efg. \tag{III-7}$$

III - 2. Fixed effects case

In the fixed effects case it is assumed that the ϵ_{ijh} is a random variate independent of the other effects and that the following relations hold:

$$\sum_{i=1}^{p} \alpha_i = 0 \quad ; \tag{III-8}$$

$$\sum_{\substack{i=1 \\ \neq j}}^{p} \delta_{ij} = 0 = \sum_{\substack{j=1 \\ \neq i}}^{p} \delta_{ij} \quad ; \tag{III-9}$$

$$E[\alpha_i^2] = \alpha_i^2 \quad ; \tag{III-10}$$

$$E[\mu^2] = \mu^2 \quad ; \tag{III-11}$$

$$E[\delta_{ij}^2] = \delta_{ij}^2 \quad ; \tag{III-12}$$

$$E[\epsilon_{ijh}^2] = \sigma_\epsilon^2 \quad ; \tag{III-13}$$

$$E[\epsilon_{ijh}\,\epsilon_{efg}] = 0 \quad \text{except when } ijh = efg; \tag{III-14}$$

$$E[\alpha_i] = \alpha_i \quad ; \tag{III-15}$$

$$E[\alpha_i\mu] = \alpha_i\mu \quad ; \tag{III-16}$$

$$E[\alpha_i \delta_{ij}] = \alpha_i \delta_{ij} \quad ; \tag{III-17}$$

$$E[\delta_{ij}\mu] = \delta_{ij}\mu \quad . \tag{III-18}$$

With the above assumptions the expected values of G' and S' are, respectively:

$$E[G'] = \frac{1}{np(p-1)(p-2)} \left\{ p \sum_i E[X_{i..}^2] - 4E[X_{...}^2] \right\}$$

$$= \frac{1}{np(p-1)(p-2)} \left\{ p \sum_i [(n(p-1)\mu + n(p-2)\alpha_i)^2 + n(p-1)\sigma_\epsilon^2] \right.$$

$$\left. - 4 \left[\frac{n^2p^2(p-1)^2}{4} \mu^2 + \frac{np(p-1)}{2} \sigma_\epsilon^2 \right] \right\}$$

$$= \frac{1}{np(p-1)(p-2)} \left\{ np(p-1)(p-2)\sigma_\epsilon^2 + pn^2(p-2)^2 \sum \alpha_i^2 \right\}$$

$$= \sigma_\epsilon^2 + \frac{n(p-2)}{p-1} \sum \alpha_i^2 \quad ; \tag{III-19}$$

$$E[S'] = \frac{2}{np(p-3)} \left\{ \sum_{i<j} E[Y_{ij.}^2] - \frac{\sum E[Y_{i..}^2]}{(p-2)} \right.$$

$$\left. + 2E[Y_{...}^2]/(p-1)(p-2) \right\}$$

$$= \frac{2}{np(p-2)(p-3)} \left\{ (p-2) \sum_{i<j} [n^2(\mu+\alpha_i+\alpha_j+\delta_{ij})^2 + \frac{np(p-1)}{2} \sigma_\epsilon^2] \right.$$

$$- [\sum_i (n(p-1)\mu + n(p-2)\alpha_i)^2 + np(p-1)\sigma_\epsilon^2]$$

$$\left. + \frac{2}{(p-1)} \left[\frac{n^2p^2(p-1)^2}{4} \mu^2 + \frac{np(p-1)}{2} \sigma_\epsilon^2 \right] \right\}$$

$$= \frac{2}{np(p-2)(p-3)} \left\{ n^2(p-2) \sum_{i<j} (\alpha_i+\alpha_j)^2 - n^2(p-2)^2 \sum \alpha_i^2 \right.$$

$$+ \sigma_\epsilon^2 \left(\frac{np(p-1)(p-2)}{2} - np(p-1) + np \right)$$

$$+ (p-2)n^2 \sum_{i<j} \delta_{ij}^2 \Big\}$$

$$= \sigma_\epsilon^2 + \frac{2n}{p(p-3)} \sum_{i<j} \delta_{ij}^2 \quad . \tag{III-20}$$

IV. Expectation of Mean Square for General Combining Ability of ith Line

IV. - 1. Random effects case

Since E $[\alpha_i^2] = \sigma_\alpha^2$ and since it is considered that α_i is a randomly selected member from an infinite population, it is inappropriate to talk about the expectation of the mean square for general combining ability for the ith line. Instead, one should talk about estimates of σ_α^2 obtained from a randomly selected line. The expectation of $\hat{\alpha}_i^2$ is:

$$E[\hat{\alpha}_i^2] = E\Big[\Big\{ \big(Y_{i..} - \frac{2Y_{...}}{p}\big) \Big/ n(p-2) \Big\}^2\Big]$$

$$= \frac{1}{n^2(p-2)^2} E\Big[\Big\{ n(p-1)(\mu+\alpha_i) + n \sum_{\substack{j=1 \\ j\neq i}}^{p} (\alpha_j+\delta_{ij}) + \sum_{j\neq i} \sum_{h} \epsilon_{ijh}$$

$$- \frac{2}{p}\Big(\frac{np(p-1)}{2} \mu + n(p-1) \sum_{i=1}^{p} \alpha_i + n \sum_{i<j=2}^{p} \delta_{ij} + \sum_{i<j=2}^{p} \sum_{h=1}^{n} \epsilon_{ijh} \Big) \Big\}^2\Big]$$

$$= \frac{1}{n^2(p-2)^2} E\Big[\Big\{ n(p-2)\alpha_i + \Sigma\alpha_i\big(n - \frac{2n(p-1)}{p}\big) + n \sum_{j\neq i} \delta_{ij}$$

$$+ \sum_{j\neq i} \sum_{h} \epsilon_{ijh} - \frac{2n}{p} \sum_{i<j} \delta_{ij} - \frac{2}{p} \sum_{i<j} \sum_{h} \epsilon_{ijh} \Big\}^2 \Big]$$

$$= \frac{1}{n^2(p-2)^2} \Big[\frac{n^2(p-1)(p-2)^2}{p}\sigma_\alpha^2 + \frac{n^2(p-1)(p-2)}{p}\sigma_\delta^2$$

$$+ \frac{n(p-1)(p-2)}{p} \sigma_\epsilon^2$$

$$= \frac{p-1}{p}\sigma_\alpha^2 + \frac{p-1}{p(p-2)}\sigma_\delta^2 + \frac{p-1}{np(p-2)}\sigma_\epsilon^2$$

$$= \frac{(p-1)}{np(p-2)}\left\{ \sigma_\epsilon^2 + n\delta_\delta^2 + n(p-2)\,\sigma_\alpha^2 \right\} . \qquad \text{(IV-1)}$$

Therefore, the estimated component of variance from the ith for general combining ability is:

$$\hat{\sigma}_\alpha^2 = p\left\{Y_{i..} - 2Y_{...}/p\right\}^2 / n^2(p-1)(p-2)^2 - \frac{S'}{n(p-2)} . \qquad \text{(IV-2)}$$

IV - 2. Fixed effects case

For this case it is assumed that $E[\alpha_i^2] = \alpha_i^2$, $E[\alpha_i] = \alpha_i$, $\sum_{i=1}^{p} \alpha_i$ = zero (i.e., the p α_i constitute the entire (finite) population), $\sum_{j\neq i} \delta_{ij}$ = zero, and the ϵ_{ijh} are random independent variates with mean zero and variance σ_ϵ^2. Thus,

$$E[\hat{\alpha}_i^2] = \frac{1}{n^2(p-2)^2} E\left[\left\{Y_{i..} - 2Y_{...}/p\right\}^2\right]$$

$$= \frac{1}{n^2(p-2)^2} E\left[\left\{ n(p-1)\mu + n(p-2)\alpha_i + \sum_{j\neq i}\sum_h \epsilon_{ijh} - \frac{2}{p}\left(\frac{np(p-1)\mu}{2} + \sum_{i<j}\sum_h \epsilon_{ijh}\right)\right\}^2\right]$$

$$= \frac{1}{n^2(p-2)^2}\left[n^2(p-2)^2\alpha_i^2 + \frac{n(p-1)(p-2)}{p}\sigma_\epsilon^2 \right]$$

$$= \alpha_i^2 + \frac{p-1}{np(p-2)}\sigma_\epsilon^2 ; \qquad \text{(IV-3)}$$

which is what was obtained by Sprague and Tatum [7] and Federer [2]. Therefore, α_i^2 is estimated as follows:

$$\hat{\alpha}_i^2 = \frac{1}{n^2(p-2)^2}(Y_{i..} - 2Y_{...}/p)^2 - \frac{np(p-2)}{p-1} E$$

$$= \frac{p-1}{np(p-2)}\left\{ \frac{p}{n(p-1)(p-2)}\left(Y_{i..} - 2Y_{...}/p\right)^2 - E \right\} , \qquad \text{(IV-4)}$$

which is a form of the formulae obtained by Sprague and Tatum [7] and Federer [2] .

It should be pointed out that $\hat{\alpha}_i^2$ is not a variance component in the usual sense, but is merely the square of an estimate of a parameter. The usefulness of a quantity like $\hat{\alpha}_i^2$ is not known since the squaring obliterates the algebraic sign. The quantity $\hat{\alpha}_i$ is, however, of use in estimating the general combining ability for a given line in relation to the other lines in the experiment.

With regard to either the random effects or the fixed effects case, there appears to be no difficulties in estimating and in using the general combining ability for the ith line. The nature of the experimental material and conditions will determine which model is appropriate.

IV - 3. Average over all lines

From formulae (III-1) and (III-2) we note that the estimate of σ_α^2 for the random effects case may be obtained as $\hat{\sigma}_\alpha^2 = \frac{1}{n(p-2)}\left\{ G' - S' \right\}$. (IV-5)

It is of interest to see the relationship between the p $\hat{\sigma}_\alpha^2$'s obtained from formula (IV-2) for all of the p lines and $\hat{\sigma}_\alpha^2$ obtained in (IV-5). Summing over i and dividing by p in (IV-2) we obtain

$$\hat{\sigma}_\alpha^2 = \frac{1}{n^2(p-1)(p-2)^2} \sum_{i=1}^{p} (Y_{i..} - 2Y_{...}/p)^2 - \frac{1}{n(p-2)} S'$$

$$= \frac{1}{n(p-2)} \left\{ G' - S' \right\} \qquad \text{(IV-6)}$$

which is identical to (IV-5).

For the fixed effects case, the estimate of $\Sigma\alpha_i^2$ from formula (III-19) is $\Sigma\hat{\alpha}_i^2 = \frac{p-1}{n(p-2)} \left\{ G' - E' \right\}$. (IV-7)

Now if we sum over the α_i^2 obtained from formula (IV-4) we obtain

$$\Sigma \hat{\alpha}_i^2 = \frac{1}{n^2(p-2)^2} \Sigma (Y_{i..} - 2Y_{...}/p)^2 - \frac{(p-1)}{n(p-2)} E'$$

$$= \frac{p-1}{n(p-2)} \left\{ G' - E' \right\} \qquad \text{(IV-8)}$$

which is formula (IV-7).

Thus, for both cases the average of the estimates from each line is identical with that obtained from all lines taken collectively.

V. Expectations for Specific Combining Ability for the ith Line

V - 1. Case I

For the first situation the following conditions are assumed:

(i) $E_i [\delta_{ij}^2] = \sigma_{\delta i}^2$;

(ii) $E_i [\epsilon_{ijh}^2] = \sigma_\epsilon^2$;

(iii) $E_i [\delta_{ij} \delta_{if} \text{ for } j \neq f] = 0$;

(iv) ϵ_{ijh} are random independent variates with mean zero;

(v) the p-1 lines crossed with line i represent a random sample of lines from an infinite population.

In (i) to (iii) the conditional expectations are given for the ith line (see [8] on conditional distributions and on conditional probabilities). The expectation of the mean square for specific combining ability for the ith line would not depend upon which line is being considered if we make assumption that $\sigma_{\delta i}^2 = \sigma_\delta^2$ a constant. Since this assumption was not made it is necessary to use conditional expectations.